M.E. COMPUTER SCIENCE AND ENGINEERING FIRST YEAR FIRST SEMESTER - 2024

MACHINE LEARNING

Time: 3 hours   Full Marks: 100

Answer Question N**o. 1 (Compulsory)** and any **Four** questions from the rest

1. State whether the following statements are **True** or **False**. Give reasons in favour of your answer

    a. In decision tree learning, the goal is to minimize information gain.
    b. Curse of dimensionality only makes KNN slow, but it has no effect on its accuracy.
    c. A Logistic Regression Model is more robust than the threshold classifier
    d. Support vectors are the training instances for which Lagrange multipliers are zero
    e. A dendrogram obtained by a Hierarchical clustering algorithm can be used to detect outliers

    $5 \times (1+3) = $ **20 marks**

2. a) Given the following short movie reviews, each labeled with a genre, either **comedy** or **action**:

Text words	Class
fun, couple, love, love	comedy
fast, furious, shoot	action
couple, fly, fast, fun, fun	comedy
furious, shoot, shoot, fun	action
fly, fast, shoot, love	action

    Using the above training dataset, compute the most likely class for a new document
    D: < fast, couple, shoot, fly>. Assume a Naive Bayes classifier and use add-1 smoothing for the likelihood
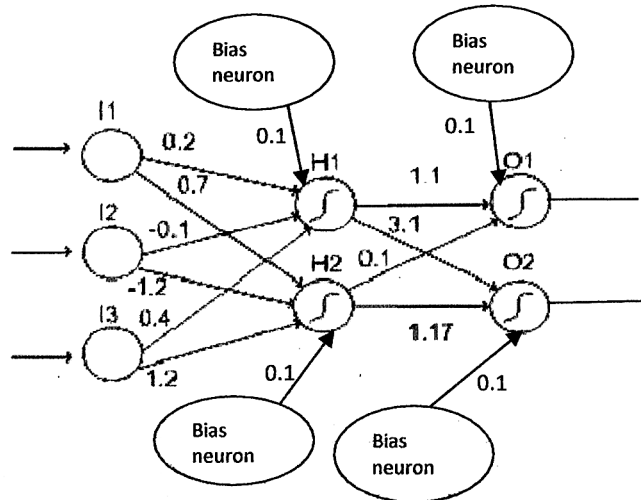
    b) Consider the following set of training examples. Use ID3 algorithm to learn a decision tree from these data (Show all necessary computations)

Instance	A1	A2	A3	Classification
1	1	0	0	Yes
2	1	0	1	Yes
3	0	1	0	No
4	1	1	1	Yes
5	1	1	0	No

    10+10=20 marks

3. a) A deep neural network is nothing but a neural network. If it is, why did deep learning not take off last time around in the 1990s?

    b) The following is an Artificial Neural Networks, with sigmoid units in the hidden layer and the output layers. The weights have been set arbitrarily between all the units.

Consider that the following training example is submitted to the net (shown in the above figure).

X1	X2	X3	O1	O2
0.2	0.5	0.3	0.9	0.1

Now show the forward pass to compute responses at H1, H2 and O1, O2 Then use backpropagation algorithm to find updates for weights associated with the connections between hidden and output layers. Assume the learning rate $\eta = 0.2$. You may assume the values of the other parameters if necessary.   5 + 15 = **20 marks**

4. a) For linear SVM, to find the maximum margin hyperplane $(w^T x + w_0 = 0)$, we need to find w and w_0 by solving the following quadratic optimization problem.:

    $min \frac{1}{2}||w||^2$ subject to $r^t(w^t x^t + w_0) \geq +1, \forall t$, where (x^t, r^t) is the t-th training instance and (w, w_0) are model parameters.

    Describe how the above quadratic programming problem is solved using Lagrange multiplies for finding the maximum margin hyperplane.

    b) Consider the one-dimensional dataset given below and classify the data point x=5.0 according to 1-, 3-, 5-, and 9-nearest neighbours. What can you observe from the results? Explain.

X	0.5	3.0	4.5	4.6	4.9	5.2	5.3	5.5	7.0	9.5
Y	NO	NO	YES	YES	YES	NO	NO	YES	NO	NO

                                                    10 + 10 = **20 marks**

5. a) Derive the weight upade rules for the logistic regression model. Consider MSE loss function.
    b) What is overfitting? How can it be dealt with in decision tree learning?
    c) What is ensembling? What are reasons for doing it? Can ensembling always produce better accuracy than the individual base learner? -explain with justifications

    $$10 + 5 + 5 = \textbf{20 marks}$$

6. a) Describe the K-means clustering algorithm with a suitable example. What is its convergence criteria? Give examples to illustrate when it may fail and why.
    b) Describe the following hierarchical clustering algorithms
       i. Single linkage   ii. Average Linkage
    c) Give an example of application where reinforcement learning can be used. Give reasons.

    $$(5+2+3)+5+5 = \textbf{20 marks}$$

7. a) What is conditional independence assumption? Why is it considered for Naive Bayes Classifier design?
    b) How continuous attributes are handled in decision tree leaning? Illustrate with an example.
    c) What is the basic difference between traditional programming and machine learning? Explain with examples
    d) With a suitable example of multi-class(>2 classes) classification output, show how the following evaluation measures are computed?
              (i) Macro F1-score (ii) Weighted F1-score

    $$5 \times 4 = \textbf{20 marks}$$